REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-01-0188

1. REPORT DATE *(DD-MM-YYYY)*
07–2003

2. REPORT TYPE
Technical

3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE

REFLECTIONS ON LOGIC & PROBABILITY IN THE CONTEXT OF CONDITIONALS

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
0601152N

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHORS

P. G. Calabrese

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

SSC San Diego
San Diego, CA 92152-5001

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
800 North Quincy Street
Arlington, VA 22217–5000

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT

20090803062

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This paper discusses various controversies surrounding the meaning and use of such conditionals as "A given B" or "If B then A" including that such Boolean fractions (1) can non-trivially carry the standard conditional probability, (2) are truth functional but with three rather than two truth values, (3) are logically and probabilistically non-monotonic, (4) can be combined with operations that extend the standard Boolean operations, and (5) allow definitions that extend Boolean deduction but do not serve as deductions themselves thereby avoiding the so-called paradoxes identified by E. Adams. A new theory of deduction with uncertain conditionals is defined in terms of the new operations on conditionals by extending the familiar equations that define deduction between Boolean propositions. This leads to several plausible forms of deduction between conditionals. These different deductive relations on conditionals give rise to different sets of implications. Methods to determine the implications of one or more conditionals with respect to the various different deductive relations are described. Three examples of deduction with uncertain conditionals are extensively examined and solved. An example about an absent-minded coffee drinker contains two so-called subjunctive or counter-factual conditionals, which pose no additional difficulty. The issue of practical computation with conditionals is addressed and the use of information entropy to cut through complexity is discussed and illustrated. Lastly there is the question of how much confidence can be attached to a probability distribution having maximum entropy. In this regard, the results of E. Jaynes concerning the concentration of distributions at maximum entropy are described along with two other theoretical approaches to this problem.

Published in *Proceedings of the Conditionals, Information, and Inference Workshop*, Hagen, Germany, G. Kern-Isberner & W. Rödder, eds., May 13-15, 2002, 27-45.

15. SUBJECT TERMS

Mission Area: Information Science

condition	deduction	uncertainty	three-valued	complexity
logic	probability	implication	entropy	Boolean fractions

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
U	U	U	UU	18	P. G. Calabrese

19B. TELEPHONE NUMBER *(Include area code)*
(619) 553-3680

REFLECTIONS ON LOGIC & PROBABILITY IN THE CONTEXT OF CONDITIONALS

Philip G. Calabrese
calabres@spawar.navy.mil

Joint & National Systems Division (Code 2737)
Space & Naval Warfare Systems Center
San Diego, CA 92152-5001

Keywords: conditional, deduction, uncertainty, three-valued, logic, probability, implication, entropy, complexity, Boolean fractions

Abstract. This paper discusses various controversies surrounding the meaning and use of such conditionals as "A given B" or "If B then A" including that such Boolean fractions 1) can non-trivially carry the standard conditional probability, 2) are truth functional but with three rather than two truth values, 3) are logically and probabilistically non-monotonic, 4) can be combined with operations that extend the standard Boolean operations, and 5) allow definitions that extend Boolean deduction but do not serve as deductions themselves thereby avoiding the so-called paradoxes identified by E. Adams. A new theory of deduction with uncertain conditionals is defined in terms of the new operations on conditionals by extending the familiar equations that define deduction between Boolean propositions. This leads to several plausible forms of deduction between conditionals. These different deductive relations on conditionals give rise to different sets of implications. Methods to determine the implications of one or more conditionals with respect to the various different deductive relations are described. Three examples of deduction with uncertain conditionals are extensively examined and solved. An example about an absent-minded coffee drinker contains two so-called subjunctive or counterfactual conditionals, which pose no additional difficulty. The issue of practical computation with conditionals is addressed and the use of information entropy to cut through complexity is discussed and illustrated. Lastly there is the question of how much confidence can be attached to a probability distribution having maximum entropy. In this regard the results of E. Jaynes concerning the concentration of distributions at maximum entropy are described along with two other theoretical approaches to this problem.

1. Introduction. Thirty–five years ago the theories of logic and of probability were conspicuously unfinished, missing a division operation to represent conditional statements. Even today many people still reduce all conditional statements such as "if B then A" to the unconditioned (or universally conditioned) statement "A or not B", the so-called "material conditional", even though it has long been recognized that the material conditional is of no use in estimating the probability of A in the context of the truth of B. The latter probability is the well-known conditional probability of A given B, the ratio of the probability of both A and B to the probability of B. The conditional probability is never greater than, and is generally much less than, the probability that A is true or B is false. Only when B is certain or when A is certain given the truth of B do the two expressions yield essentially the same result. Even when B is false, they differ since the ratio is undefined while the material conditional has probability 1. This has all been quantified for instance in [Cal87]. Yet for purposes of doing 2-valued logic, the material conditional works just fine. Mathematicians have long proved their theorems of the form "if B then A" by proving that in all cases either A is true or B is false.

However when B is uncertain or when A is uncertain given the truth of B, the material conditional is not an appropriate Boolean proposition to represent the conditional statement "if B then A". Nor is there any other Boolean proposition that can serve the purpose of both logic and probability as early shown by D. Lewis [Lew75]. This non-existence is reminiscent of results throughout the history of mathematics that preceded the invention of new numbers needed to satisfy some relationships that naturally arose. The irrational numbers were needed to represent the length of the hypotenuse of a square in terms of the length of a side of that square; complex numbers were invented to solve polynomial equations such as $x^2 + 1 = 0$ and integer fractions were invented to have numbers that could solve equations like $3\,x = 20$. In each case, mathematicians didn't stop with the declaration that there were no such numbers in the existing system; they instead invented new numbers that included the old ones but also solved the desired equations. The same thing has worked in the case of events and propositions [Cal87, Cal94] and the result is no less profound. The more surprising thing is that it has taken so long for the development to occur in the case of events and propositions. Apart from Boole himself, such a system of ordered pairs was envisioned by a few researchers including G. Schay [Sch68] and Z. Domotor Dom69], but these developments didn't go far enough in the right direction before getting bogged down. It is now clear, however, that a system of *ordered pairs* of probabilistic events or of logical propositions can be defined to represent conditional statements, avoid the triviality results of Lewis [Lew75], and be assigned the standard conditional probability.

These operations, on the ordered pairs of events or propositions (A|B), "A given B", have been extensively analyzed and motivated in [Cal87, Cal94, Cal02]. Using ' to denote "not" and juxtaposition to denote "and" these operations on conditionals are:

a) $(A|B)' = (A' \mid B)$.

That is, "not (A given B)" is equivalent to "(not A) given B".

b) (A|B) or (C|D) = ((AB or CD) | (B or D))

The right hand side is "given either conditional is applicable, at least one is true".

c) (A|B) and (C|D) = [ABD' or ABCD or B'CD] | (B or D)

The right hand side is "given either conditional is applicable, at least one is true while the other is not false". It can be rewritten as [AB(CD or D') or CD(AB or B') | B or D)].

d) $(A|B) \mid (C|D) = (A \mid (B)(C|D))$

The right hand side is "given B and (C|D) are *not false,* A is true."

By writing B as a conditional $(B \mid \Omega)$ with the universe Ω as condition the conjunction (B)(C|D) in d) reduces to $B(C \vee D')$ using operation c).

This system of "Boolean fractions" $(\mathcal{B}|\mathcal{B})$ includes the original events or propositions $\mathcal{B}$ as a subsystem and also satisfies the essential needs of both logic and conditional probability. Two conditionals (A|B) and (C|D) are *equivalent* (=) if and only if B=D and AB = CD. As with the past extensions of existing number systems, some properties no longer hold in the new system. For instance, the new system is not wholly distributive as are Boolean propositions.

As with any new system of numbers there has been quite a lot of resistance to this new algebra of conditionals. Some researchers (see [Goo91A, Hai96]), recognizing the virtue of a system of ordered pairs of events to represent conditional events, have nevertheless disputed the choice of extended operations on those ordered pairs. However, the operations for "or" and

"and" in [Cal87] were independent rediscoveries of the two so-called "quasi" operations for "or" and "and" early employed by E. Adams [Ada66, Ada86], a pioneer researcher of conditionals writing in the philosophical literature. Adams calls these operations "quasi" merely because they are not "monotonic". That is, combining two conditionals with "and" does not always result in a new conditional that implies each of the component conditionals. Nor does combining two conditionals with "or" always result in a conditional that is implied by each of the component conditionals. This seems rather counter intuitive when considered in the abstract because we are all so imbued with equal-condition thinking. But when two conditionals with different conditions are combined as in operations b) or c), the result is a conditional whose condition is the disjunction ("or") of the two original conditions. By expanding the context in this way probabilities have more freedom to change up or down. Deduction is also much more complicated when dealing with conditionals with different conditions, but now a successful extension of Boolean deduction for uncertain conditionals has been developed [Cal90, Cal91, Cal02].

Another issue that arises with conditionals is their truth functionality. Are conditionals "true" or "false" like ordinary propositions or events? Even the ancient Greeks were troubled by this question. For some reason Adams seems to take the attitude [Ada98, p.65, footnote] that "inapplicable" is not really a 3rd truth-value that can be assigned to a conditional. On the other hand, B. De Finetti [DeF36] early asserted that a conditional has three, rather than two, truth-values: If the condition B is true, then "A given B" is true or false depending on the truth of A. But when B is false, De Finetti asserted that the conditional was neither true nor false, but instead required a third truth-value, which he unfortunately identified with "unknown" and therefore assigned a numerical value somewhere between 0 and 1. But a conditional with a false condition is not "unknown"; it is "inapplicable". For instance, if I am asked, "if you had military service, in which branch did you serve?" I don't answer "unknown". I answer "inapplicable" because I haven't had military service. The question and its answer are not assigned a truth-value between 0 and 1; they are essentially ignored. The answer "unknown" would be appropriate by someone who thought I had military service but did not know in which branch I served.

While it is not immediately obvious, the question of what operations are used to combine conditional propositions is essentially equivalent to the question of which of the three truth-values should be assigned to the nine combinations of the truth (T), falsity (F) or inapplicability (I) for two different conditionals. See [Cal93, p.7] for a proof. This approach was taken by A. Walker [Wal94] to determine those few operations on conditionals that satisfy natural requirements such as being commutative and idempotent. This approach was also employed in [Cal02] to provide careful motivations and a complete characterization of the 4 operations on conditionals a) – d) listed above and originally grouped together in [Cal87]. Three of these operations in the form of 3-valued truth tables were identified by B. Sobocinski [Sob52, Res69], but his 4^{th} operation was very different from the operation d) in [Cal87]. Similarly, Adams easily identified the negation operation for conditionals, but passed over the 4^{th} iterated conditioning operation employed here because he interprets a conditional as an implication instead of as a new object - an event or proposition in a given context.

Recently, Adams reconsidered the issue of "embedded" or iterated conditionals [Ada98, p.268] and the so-called "import-export" principle which asserts that $((A \mid B) \mid C) = (A \mid B \text{ and } C)$ for any expressions A, B and C. Operation d) is a restricted form of this principle, which can be used to reduce any iterated conditional to a simple conditional with Boolean components. For propositions A, B, C, D, E, and F, a more general form of the import-export law follows from operations a) - d):

$$[(A|B) \mid (C|D)] \mid (E|F) = (A|B) \mid [(C|D)\,(E|F)] \tag{1.1}$$

Using "import-export" Adams sites the following example as a counter example of the basic logical principle of modus ponens that A is always a logical consequence of B and (A|B). Noting that by import-export, ((HD | H) | D) = (HD | HD), and that the latter is a logical necessity, Adams gives the example

$$D \text{ and } (\,(HD \mid H) \mid D\,) \text{ implies } (HD \mid H), \tag{1.2}$$

Which, according to Adams, should be valid by modus ponens. For instance, interpreting D as "it is a dog" and H as "it is heavy (500 pounds)" modus ponens seems to fail because the implication (HD | H), that "it is a heavy dog given that it is heavy" should not logically follow from D and "(HD | H) given D". Adams mentions three authors who each take a different direction here, one accepting "import-export", one accepting modus ponens, and one accepting both with reservations about modus ponens.

But the difficulties raised by this example disappear when it is remembered that with modus ponens, it is not just "A" that is a logical consequence of "B and (A|B)", but rather "A and B" that is the logical consequence. And since conditionals are not logically monotonic, "A and B" does not necessarily imply "A" alone, as Adams has elsewhere shown. For conditionals, "A and B" may no longer imply "A" and may also have larger probability than "A" alone.

Therefore, the logical implication of the left side of equation 1.2 is "D and (HD | H)", which by operation c) reduces to just D, and D is certainly a valid implication of the left side of 1.2. So the "paradox" arises because the notion that "B and A" must logically imply B is false for conditionals.

For example, consider a single roll of a fair die with faces numbered 1 through 6. The conditional (2 | even) representing "2 comes up given an even number comes up" has conditional probability 1/3, and it surely logically implies itself by any intuitive concept of implication. Now conjoin the conditional (1 or 3 | < 5), representing "1 or 3 comes up given the roll is less than 5", with (2 | even) and the result by operation c) is (1 or 3 | not 5), which obviously does not logically imply (2 | even) by any intuitive concept of logical implication. Note also that (1 or 3 | not 5) has conditional probability 2/5, which is larger than 1/3, the conditional probability of (2 | even). All of these situations have been analyzed in [Cal02]. Adams gives a similar example [Ada98, p. 273] that can be handled in the same way.

Concerning embedded conditionals, Adams claims [Ada98, p. 274] that, "So far no one has come up with a pragmatics that corresponds to the truth-conditional or probabilistic semantics of the theories that they prose …". However Adams has too quickly passed over the 4-operation system of Boolean fractions (conditionals events) recounted here, and he has not yet examined the additional theory of deduction defined in terms of the operations on those conditionals.

To repeat, most if not all of these so-called paradoxes of embedded conditionals and logical deduction arise from the unwarranted identification of the conditional (A|B) with the logical implication of A by B. Others arise by forgetting that conditionals are logically non-monotonic. However, when (A|B) is taken as a new object and deduction is defined in terms of the operations a) – d), these paradoxes disappear. Just as it is in general impossible to force Boolean propositions to carry the conditional probability, so too is it impossible to force conditionals to serve as implication relations. The latter must be separately defined in terms of, or at least consistent with, the chosen operations on conditionals.

In Section 2.1 and 2.2 the essentials of the theory of deduction with uncertain conditionals are recounted including some refinements such as Definition 2.2.4 of the "conjunction property". Section 2.3 provides three new illustrative examples of deduction with uncertain conditionals. Section 2.3.1 addresses the familiar question of what can be deduced by transitivity with conditionals. That is, what can be deduced from "A given B" and "B given C"? Section 2.3.2 analyzes a set of three rather convoluted conditionals concerning an absent-minded coffee drinker. Two of the three conditionals are so-called non-indicative, also called subjunctive or counterfactual conditionals. Such conditionals seem to pose no additional difficulty for this theory of deduction. In Section 2.3.3 the absent-minded coffee drinker example is modified to make it a valid deduction in the two-valued Boolean logic. The implications with respect to various deductive relations are again determined. Section 3 addresses the issue of practical computation of combinations of conditionals and deductions with conditionals. Section 3.1 illustrates the difficulties and complexities of pure Bayesian analysis when applied to the "transitivity example" of Section 2.3.1. Section 3.2 discusses the use of entropy in information processing as a reasonable and principled way to cut through complexity and solve for unknown probabilities and conditional probabilities. This idea has already been successfully implemented in the computer program SPIRIT developed at Hagen University by a team headed by W. Rödder. Section 3.3 addresses the question of the confidence that can be attached to probabilities determined by the maximum entropy solution. In this regard the separate work of E.T. Jaynes, S. Amari, and A. Caticha are described, especially that of Jaynes, who proves an entropy concentration theorem that provides a statistical measure of the fraction of eligible probability distributions whose entropy falls below a specified critical value.

2. Deduction with Uncertain Conditionals. Deduction for uncertain conditionals must be defined in terms of the operations a) – d) on conditionals listed in the introduction. For instance, if (A|B) and (C|D) are two conditionals, we may wish to define deduction of (C|D) by (A|B) to mean that the conjunction (A|B) (C|D) of the two conditionals should be equivalent to (A|B) as is the case with Boolean propositions. Recall that for Boolean propositions p implies q can be defined with the conjunction operation by the equation "p and q = p". Alternately, we could use the disjunction operation and define this same implication as "p or q = q". Still other ways exist such as "q or not p = 1 (true)". Surprisingly, in the realm of conditionals none of these definitions of implication are equivalent to one another! This has all been extensively developed in [Cal90, Cal91, Cal94] and especially [Cal02]. This development will be summarized and streamlined in sections 2.1 and 2.2.

2.1 Deductive Relations. The expression "B ≤ A" is used to signify "B implies A" because for Boolean propositions this implication is equivalent to saying that "the instances of B are a subset of the instances of A". This is also the appropriate interpretation in case that A and B are probabilistic events. Some readers may wish to mentally substitute the entailment arrow ⇒ for ≤ to connote deduction.

Definition 2.1.1. An implication or deductive relation, ≤, on conditionals is a reflexive and transitive relation on the set of conditionals.

For instance, one such deductive relation is $\leq_{bo}$:

$$(A|B) \leq_{bo} (C|D) \qquad \text{if and only if} \qquad B = D \text{ and } AB \_ CD \qquad (2.1.1)$$

That is, conditional (A|B) implies conditional (C|D) with respect to this deductive relation if and only if the conditions B and D are equivalent propositions or events, and within this common condition, proposition A implies proposition C. This is called *Boolean deduction* because it is just ordinary Boolean deduction when applied to conditionals with the same condition, and a conditional can only imply another conditional provided they have equivalent conditions.

Using conjunction ($\wedge$) to define implication yields *Conjunctive Implication* ($\leq_\wedge$):

$$(A|B) \leq_\wedge (C|D) \quad \text{if \& only if} \quad (A|B) \wedge (C|D) = (A|B) \qquad (2.1.2)$$

For $\leq_\wedge$ the conjunction of two or more conditionals always implies each of its components.

Using disjunction ($\vee$) to define implication yields *Disjunctive Implication* ($\leq_\vee$):

$$(A|B) \leq_\vee (C|D) \quad \text{if \& only if} \quad (A|B) \vee (C|D) = (C|D) \qquad (2.1.3)$$

For $\leq_\vee$ the disjunction of two or more conditionals is always implied by each of the component conditionals.

Applying the material conditional equation "q or not p = 1" to conditionals yields what is called *Probabilistically Monotonic Implication* ($\leq_{pm}$):

$$(A|B) \leq_{pm} (C|D) \quad \text{if \& only if} \quad (C|D) \vee (A|B)' = (\Omega \mid D \vee B) \qquad (2.1.4)$$

For $\leq_{pm}$ any conditional (C|D) implied by (A|B) has conditional probability no less than P(A|B). Here, the universal proposition is denoted "1" and the universal event is Ω.

In [Cal91, Cal02] the defining equations on the right side of the definitions (2.1.1 – 2.1.4) have been reduced to Boolean deductive relations between the component Boolean propositions. For instance, 2.1.2 reduces to the two Boolean implications, ($A \vee B' \leq C \vee D'$) and ($B' \leq D'$); 2.1.3 reduces to ($AB \leq CD$) and ($B \leq D$); and 2.1.4 reduces to ($A \vee B' \leq C \vee D'$) and ($AB \leq CD$). Thus between two conditionals (A|B) and (C|D) four elementary Boolean deductive relations arise: $B \leq D$, $AB \leq CD$, $A \vee B' \leq C \vee D'$ and $B' \leq D'$. What is implied by these implication relations is applicability, truth, non-falsity and inapplicability respectively. They have been denoted $\leq_{ap}$, $\leq_{tr}$, $\leq_{nf}$, and $\leq_{ip}$ respectively where "ap" means "applicable", "tr" means "truth"", "nf" means "non-falsity" and "ip" means "inapplicable". This leads to a hierarchy of deductive relations on conditionals as one, two, three or all four of these different Boolean relations are assumed necessary for a deductive relation $(A|B) \leq_x (C|D)$ to hold between two conditionals (A|B) and (C|D). See Figure 2.1. Actually, except for $\leq_{bo}$ all of these deductive relations can be defined in terms of just one or two of the four elementary ones because, for instance, the combined properties of $\leq_{ap}$ and $\leq_{nf}$ are equivalent to those of $\leq_{m0}$. Similarly, the combined properties of $\leq_{tr}$ and $\leq_{ip}$ are equivalent to those of $\leq_{m\wedge}$.

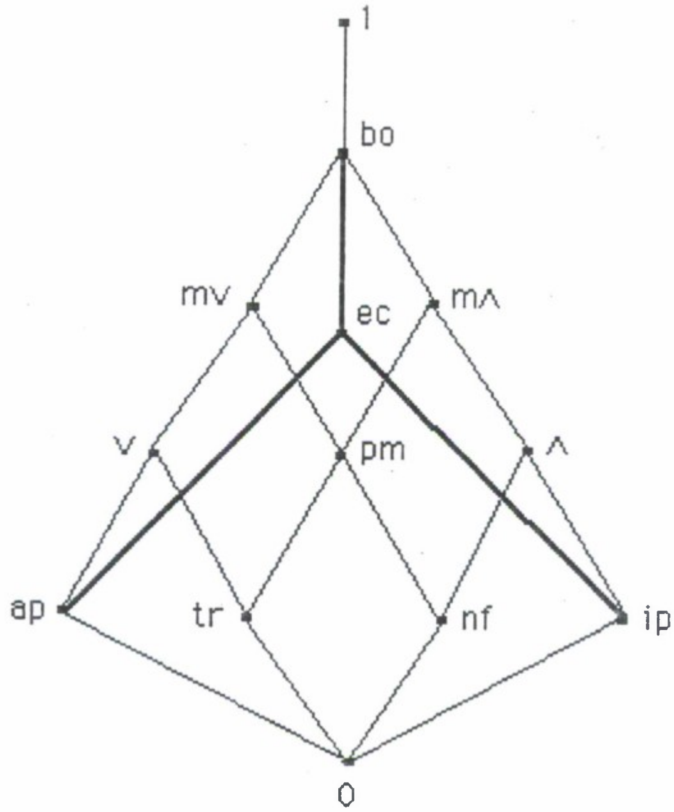

Figure 2.1 Hierarchy of Implications (Deductive Relations) for Conditionals

Trivial Implications

1 - Implication of Identity ($\leq_1$)

$(a|b) \leq_1 (c|d)$ iff $(a|b) = (c|d)$

0 - Universal Implication

$(a|b) \leq_0 (c|d)$ for all $(a|b)$ & $(c|d)$

Two Elementaries Combined

∨ - Disjunctive Implication ($\leq_\vee$)

$(a|b) \leq_\vee (c|d)$ iff $b \leq d$ and $ab \leq cd$

pm - Probabilistically Monotonic Implication; ($\leq_{pm}$)

$(a|b) \leq_{pm} (c|d)$ iff $ab \leq cd$ and $(a \vee b') \leq_{nf} (c \vee d')$

∧ - Conjunctive Implication ($\leq_\wedge$)

$(a|b) \leq_\wedge (c|d)$ iff $d \leq b$ and $(a \vee b') \leq_{nf} (c \vee d')$

ec - Implication of Equal Conditions ($\leq_{ec}$)

$(a|b) \leq_{ec} (c|d)$ iff $b = d$

Elementary Implications

tr - Implication of Truth ($\leq_{tr}$)

$(a|b) \leq_{tr} (c|d)$ iff $ab \leq cd$

nf - Implication of Non-Falsity ($\leq_{nf}$)

$(a|b) \leq_{nf} (c|d)$ iff $(a \vee b') \leq_{nf} (c \vee d')$

ap - Implication of Applicability ($\leq_{ap}$)

$(a|b) \leq_{ap} (c|d)$ iff $b \leq d$

ip - Implication of Inapplicability ($\leq_{ip}$)

$(a|b) \leq_{ip} (c|d)$ iff $d \leq b$

Three Elementaries Combined

m∨ - (Probabilistically) Monotonic and Applicability Implication ($\leq_{m\square}$)

m∧ - (Probabilistically) Monotonic and Inapplicability Implication ($\leq_{m\square}$)

Four Elementaries Combined

bo - Boolean Deduction ($\mathcal{B}$ | fixed b) ($\leq_{bo}$)

$(a|b) \leq_{bo} (c|d)$ iff $b = d$ and $ab \leq cd$

2.2 Deductively Closed Sets of Conditionals. Having defined the idea of a deductive relation on conditionals it is now possible to define the set of implications of a set of conditionals with respect to such a deductive relation.

Definition 2.2.1. A subset $\mathcal{H}$ of conditionals is said to be a *deductively closed set* (DCS) with respect to a deductive relation $\leq_x$ if and only if $\mathcal{H}$ has both of the following properties:

If $(A|B) \in \mathcal{H}$ and $(C|D) \in \mathcal{H}$ then $(A|B) \wedge (C|D) \in \mathcal{H}$, and

If $(A|B) \in \mathcal{H}$ and $(A|B) \leq_x (C|D)$ then $(C|D) \in \mathcal{H}$.

A set of conditionals with the first property is said to have the *conjunction property* and a set of conditionals satisfying the second property is said to have the *deduction property*.

The following theorem states that the intersection of two DCS's with respect to two different deductive relations is a DCS with respect to the deductive relation formed by combining the requirements of those two deductive relations.

Theorem 2.2.2. Conjunction Theorem for Deductively Closed Sets with respect to two Deductive Relations. If $\mathcal{H}_x$ is a deductively closed set of conditionals with respect to a deductive relation $\leq_x$, and $\mathcal{H}_y$ is a deductively closed set of conditionals with respect to a deductive relation $\leq_y$, then the intersection $\mathcal{H}_x \cap \mathcal{H}_y$ is a DCS, $\mathcal{H}_{x \cup y}$, with respect to the *combined deductive relation* $\leq_{x \cap y}$ defined by:

$$(A|B) \leq_{x \cap y} (C|D) \quad \text{if and only if} \quad (A|B) \leq_x (C|D) \quad \text{and} \quad (A|B) \leq_y (C|D).$$

The proof is very straightforward including showing that $\leq_{x \cap y}$ is a deductive relation. However, in general not all DCS's with respect to $\leq_{x \cap y}$ are intersections of DCS's with respect to the component deductive relations $\leq_x$ and $\leq_y$.

Definition 2.2.3 Deductive Implications of a set J of conditionals. If J is any subset of conditionals, $\mathcal{H}_x(J)$ will denote the smallest deductively closed subset with respect to $\leq_x$ that includes J. We say that $\mathcal{H}_x(J)$ is the deductive extension of J with respect to $\leq_x$, or that J generates or implies $\mathcal{H}_x(J)$ with respect to $\leq_x$. A DCS is *principal* if it is generated by a single conditional.

Definition 2.2.4. Conjunction Property for Deductive relations. A deductive relation $\leq_x$ has the *conjunction property* if and only if

$$(A|B) \leq_x (C|D) \text{ and } (A|B) \leq_x (E|F) \text{ implies } (A|B) \leq_x (C|D) \wedge (E|F).$$

(Note: this is different from the conjunction property satisfied by a set of conditionals.)

Theorem 2.2.5. Principal Deductively Closed Sets. With respect to any deductive relation $\leq_x$ having the conjunction property the deductively closed set generated by a single conditional (A|B) is the set of conditionals that subsume it with respect to the deductive relation. That is, $\mathcal{H}_x\{(A|B)\} = \{(Y|Z): (A|B) \leq_x (Y|Z)\}$. $\mathcal{H}_x\{(A|B)\}$ will be denoted by $\mathcal{H}_x(A|B)$.

Proof of Theorem 2.2.5. $\mathcal{H}_x(A|B)$ has the conjunction property. For suppose that (C|D) and (E|F) are in $\mathcal{H}_x(A|B)$. So $(A|B) \leq_x (C|D)$ and $(A|B) \leq_x (E|F)$. Therefore $(A|B) \leq_x (C|D)(E|F)$, by the conjunction property of $\leq_x$. So $(C|D)(E|F) \in \mathcal{H}_x(A|B)$. $\mathcal{H}_x(A|B)$ obviously also has the deduction property by the transitivity of any deductive relation $\leq_x$. Therefore $\mathcal{H}_x(A|B)$ is a DCS of conditionals. Clearly any DCS containing (A|B) must also include $\mathcal{H}_x(A|B)$. So $\mathcal{H}_x(A|B)$ is the smallest DCS containing (A|B).

Theorem 2.2.6. The four elementary deductive relations $\leq_{ap}$, $\leq_{tr}$, $\leq_{nf}$, and $\leq_{ip}$ on conditionals and their combinations, have the conjunction property of Definition 2.2.4.

Proof of Theorem 2.2.6. Suppose that (A|B) $\leq_{ap}$ (C|D) and (A|B) $\leq_{ap}$ (E|F). So B ≤ D and B ≤ F. So B ≤ (D ∧ F) ≤ (D ∨ F). Therefore (A|B) $\leq_{ap}$ (C|D) ∧ (E|F) = (CDF' ∨ D'EF ∨ CDEF | D ∨ F) because B ≤ D ∨ F. Suppose next that (A|B) $\leq_{tr}$ (C|D) and (A|B) $\leq_{tr}$ (E|F). So AB ≤ CD and AB ≤ EF. Therefore (A|B) $\leq_{tr}$ (C|D) ∧ (E|F) because AB ≤ (CD) ∧ (EF) ≤ (CDF' ∨ D'EF ∨ CDEF) ∧ (D ∨ F). Suppose next that (A|B) $\leq_{nf}$ (C|D) and (A|B) $\leq_{nf}$ (E|F). So (A ∨ B') ≤ (C ∨ D') and (A ∨ B') ≤ (E ∨ F'). Therefore (A|B) $\leq_{nf}$ (C|D) ∧ (E|F) because (A ∨ B') ≤ (C ∨ D') ∧ (E ∨ F') = (CD ∨ D') ∧ (EF ∨ F') = (CDEF ∨ D'EF ∨ CDF') ∨ D'F', which is just (CDF' ∨ D'EF ∨ CDEF) ∨ (D ∨ F)'. Fourthly, suppose that (A|B) $\leq_{ip}$ (C|D) and (A|B) $\leq_{ip}$ (E|F). So B' ≤ D' and B' ≤ F'. Therefore (A|B) $\leq_{ip}$ (C|D) ∧ (E|F) because B' ≤ D' ∧ F' = (D ∨ F)'. Finally, Suppose that (A|B) $\leq_{x \cap y}$ (C|D) and (A|B) $\leq_{x \cap y}$ (E|F) where x and y are in {ap, tr, nf, ip}. So (A|B) $\leq_x$ (C|D) and (A|B) $\leq_y$ (C|D) and (A|B) $\leq_x$ (E|F) and (A|B) $\leq_y$ (E|F). Therefore (A|B) $\leq_x$ (C|D) and (A|B) $\leq_x$ (E|F) and so (A|B) $\leq_x$ (C|D) ∧ (E|F). Similarly (A|B) $\leq_y$ (C|D) ∧ (E|F). Therefore (A|B) $\leq_{x \cap y}$ (C|D) ∧ (E|F).

Corollary 2.2.7. If $\leq_x$ is one of the elementary deductive relations $\leq_{ap}$, $\leq_{tr}$, $\leq_{nf}$, and $\leq_{ip}$ or a deductive relation combining two or more of these, then the DCS generated by (A|B) with respect to $\leq_x$ is $\mathcal{H}_x$(A|B) = {(Y|Z): (A|B) $\leq_x$ (Y|Z)}.

Proof of Corollary 2.2.7. The proof follows immediately from Theorems 2.2.5 and 2.2.6.

These results allow the principal DCS's with respect the four elementary deductive relations and their combinations to be explicitly expressed in terms of Boolean relations. See [Cal02] for details. For instance, $\mathcal{H}_{ap}$(A|B) = {(Y|Z): Y any event or proposition and Z any event or proposition with B ≤ Z} = {(Y | B ∨ Z): Y and Z any events or propositions}. For the elementary deductive relations these solutions are

$$\mathcal{H}_{ap}(A|B) = \{(Y \mid B \vee Z)\text{: any events or propositions } Y \text{ and } Z \text{ in } \mathcal{B}\} \tag{2.2.1}$$

$$\mathcal{H}_{tr}(A|B) = \{ (AB \vee Y \mid AB \vee Z)\text{: any events or propositions } Y \text{ and } Z \text{ in } \mathcal{B}\} \tag{2.2.2}$$

$$\mathcal{H}_{nf}(A|B) = \{ (AB \vee B' \vee Y \mid Z)\text{: any } Y, Z \text{ in } \mathcal{B}\} \tag{2.2.3}$$

$$\mathcal{H}_{ip}(a|b) = \{(Y \mid BZ)\text{: any } Y, Z \text{ in } \mathcal{B}\} \tag{2.2.4}$$

The following result allows the principal DCS's of the deductive relations formed by combining two or more of the elementary deductive relations to be expressed as an intersection of principal DCS's of the elementary deductive relations. This result does not extend to DCS's generated by a set of conditionals.

Theorem 2.2.8. The principal DCS $\mathcal{H}_{x \cap y}$(A|B) of a single conditional (A|B) with respect to a combination deductive relation $\leq_{x \cap y}$ is the intersection of the DCS's with respect to the component deductive relations $\leq_x$ and $\leq_y$. That is, $\mathcal{H}_{x \cap y}$(A|B) = $\mathcal{H}_x$(A|B) ∩ $\mathcal{H}_y$(A|B).

Proof of Theorem 2.2.9. $\mathcal{H}_{x \cap y}$(A|B) = {(C|D): (A|B) $\leq_{x \cap y}$ (C|D)} = {(C|D): (A|B) $\leq_x$ (C|D) and (A|B) $\leq_y$ (C|D)} = $\mathcal{H}_x$(A|B) ∩ $\mathcal{H}_y$(A|B).

Using the formulas for the principal DCS's with respect to the elementary deductive relations, the principal DCS's with respect to the combined deductive relations have been calculated in [Cal02]. For the deductive relations mentioned above, the principal DCS's are:

$$\mathcal{H}_\vee(A|B) = \{\ (AB \vee Y \mid B \vee Z)\text{: any } Y, Z \text{ in } \mathcal{B}\} \tag{2.2.5}$$

$$\mathcal{H}_{pm}(A|B) = \{\ (AB \vee B' \vee Y \mid AB \vee Z)\text{: any } Y, Z \text{ in } \mathcal{B}\} \tag{2.2.6}$$

$$\mathcal{H}_\wedge(A|B) = \{\ (AB' \vee Y \mid BZ)\text{: any } Y, Z \text{ in } \mathcal{B}\} \tag{2.2.7}$$

Having described the principal DCS's of the elementary deductive relations and their combination deductive relations, these results can be used to describe the DCS's of a set of conditionals with respect to these deductive relations.

For Boolean deduction, the implications of a finite set of propositions or events is simply the implications of the single proposition or event formed by conjoining the members of that initial finite set of conditionals. One of the counter-intuitive features of deduction with a set conditionals is the necessity of considering the deductive implications of all possible conjunctions of the members of that initial set of conditionals.

Definition 2.2.10. Conjunctive Closure of a Set of Conditionals. If J is a set of conditionals then the *conjunctive closure* C(J) of J is the set of all conjunctions of any finite subset of J.

Theorem 2.2.11. Deduction Theorem. For all the elementary deductive relations $\leq_x$ and their combinations, except for $\leq_{tr}$ and $\leq_\vee$, the DCS $\mathcal{H}_x(J)$ with respect to $\leq_x$ of a set J of conditionals is the set of all conditionals implied with respect to $\leq_x$ by some member of the conjunctive closure C(J) of J. That is,

$$\mathcal{H}_x(J) \;=\; \{\ (Y|Z)\text{: } (A|B) \leq_x (Y|Z),\ (A|B) \in C(J)\}$$

For a proof see subsection 3.4.3 of [Cal02].

Corollary 2.2.12. Under the hypotheses of the Deduction Theorem, it follows from Theorem 2.2.5. (Principal Deductively Closed Sets) that

$$\mathcal{H}_x(J) \;=\; \bigcup_{(A|B)\,\in\, C(J)} \mathcal{H}_x(A|B)$$

That is, the deductively closed set with respect to $\leq_x$ generated by a subset J of conditionals is the set of all conditionals implied with respect to $\leq_x$ by some member of the conjunctive closure C(J) of J.

For most deductive relations $\leq_x$ it is necessary in general to first determine the conjunctive closure C(J) of a finite set of conditionals J in order to determine the DCS $\mathcal{H}_x(J)$ of J. However for the non-falsity, inapplicability and conjunctive deductive relations, that is for $x \in \{nf, ip, \wedge\}$, the DCS of J is $\mathcal{H}_x(J) = \mathcal{H}_x(A|B)$, where (A|B) is the single conditional formed by conjoining all the conditionals in J.

Corollary 2.2.13. With respect to the three deductive relations $\leq_{nf}$, $\leq_{ip}$, and $\leq_{\wedge}$ the DCS of a finite set of conditionals J is principal and is generated by the single conditional formed by conjoining all the conditionals in J.

Proof of Corollary 2.2.13. Let $x \in \{nf, ip, \wedge\}$, and suppose (A|B) is the conjunction of all the conditionals in the set J of conditionals. Then $\mathcal{H}_x(J) = \mathcal{H}_x(A|B)$ because with respect to $\leq_x$, $(A|B) \leq_x (Y|Z)$ for all (Y|Z) in C(J). This follows from the fact, which is easily checked, that for these deductive relations the conjunction of two conditionals always implies each of the component conditionals.

2.3 Examples of Deduction with Uncertain Conditionals. In [Cal02] the implications of the three well known "penguin postulates" have been completely described with respect to the elementary deductive relations and their combinations. In this section two more examples will be given. First the implications of the set J of the two conditionals {(A|B), (B|C)} will be determined. Of interest is the conditional (A|C), which is easily true when the initial two conditionals are certainties, but may be false when one or the other is uncertain. We are often interested in chaining deductions and inferences in this way. What are the implications and inferences to be made from knowing "A given B" and "B given C", allowing for the lack of certainty of these conditionals?

2.3.1 Transitivity Example. Consider the set J consisting of two uncertain conditionals (A|B) and (B|C). Then the conjunctive closure C(J) = {(A|B), (B|C), (A|B)(B|C)} = {(A|B), (B|C), (AB | B ∨ C)}. For $x \in \{nf, ip, \wedge\}$, by Theorem 2.2.8 on principal DCS's, the DCS generated by J is $\mathcal{H}_x(J) = \mathcal{H}_x(AB \mid B \vee C)$. So using equations 2.2.3, 2.2.4, and 2.2.7,

$$\mathcal{H}_{ip}(J) = \{\ (Y \mid (B \vee C)\, Z\ ): \text{any } Y, Z \text{ in } \mathcal{B}\}$$

$$\mathcal{H}_{nf}(J) = \{\ (AB \vee B'C' \vee Y \mid Z): \text{any } Y, Z \text{ in } \mathcal{B}\}$$

$$\mathcal{H}_{\wedge}(J) = \{\ (AB \vee Y \mid (B \vee C)\, Z): \text{any } Y, Z \text{ in } \mathcal{B}\}$$

Notice that $(A|C) \in \mathcal{H}_{nf}(J)$ by setting Y = AB' and Z = C. In that case (AB ∨ B'C' ∨ Y | Z) = (AB ∨ AB' ∨ B'C' | C) = (A ∨ B'C' | C) = (A|C). Thus with respect to the non-falsity deductive relation $\leq_{nf}$, the conditional (A|C), as expected, is implied by (A|B) and (B|C). When (A|B) and (B|C) are non-false then so is (A|C). $\mathcal{H}_{nf}(J)$ is the set of all conditionals whose conclusion includes the truth of (A|B) and also the inapplicability of both (A|B) and (B|C). By similar arguments (A|C) is in $\mathcal{H}_{ip}(J)$ and also in $\mathcal{H}_{\wedge}(J)$.

For the elementary deductive relations $\leq_x$ or some combination of them except for $\leq_{\pi}$ and $\leq_{\vee}$, by Corollary 2.2.12 the DCS generated by J is $\mathcal{H}_x(J) = \mathcal{H}_x(A|B) \cup \mathcal{H}_x(B|C) \cup \mathcal{H}_x(AB \mid B \vee C)$.

Now let x = pm. That is, consider the deductions of J with respect to the probabilistically monotonic deductive relation $\leq_{pm}$. Since $(AB \mid B \vee C) \leq_{pm} (A|B)$, therefore $\mathcal{H}_{pm}(AB \mid B \vee C) \supseteq \mathcal{H}_{pm}(A|B)$. Thus, $\mathcal{H}_{pm}(J) = \mathcal{H}_{pm}(B|C) \cup \mathcal{H}_{pm}(AB \mid B \vee C)$. So by equation 2.2.6 $\mathcal{H}_{pm}(J) = \{BC \vee C' \vee Y \mid BC \vee Z): \text{any } Y, Z \text{ in } \mathcal{B}\} \cup \{AB \vee B'C' \vee Z \mid AB \vee Z): \text{any } Y, Z \text{ in } \mathcal{B}\}$. Note that (A|C) is not necessarily a member of $\mathcal{H}_{pm}(J)$.

Furthermore, since $\leq_{pm}$ is probabilistically monotonic, all the conditionals in $\mathcal{H}_{pm}(B|C) = \{BC \vee C' \vee Y \mid BC \vee Z): \text{any } Y, Z \text{ in } \mathcal{B}\}$ have conditional probability no less than P(B|C), and all the conditionals in $\mathcal{H}_{pm}(AB \mid B \vee C) = \{AB \vee B'C' \vee Z \mid AB \vee Z): \text{any } Y, Z \text{ in } \mathcal{B}\}$ have conditional probability no less than P(AB | B ∨ C).

2.3.2 Absent-minded Coffee Drinker Example. The second example by H. Pospesel [Pos71, p.27, #78] is a typical inference problem called the "absent-minded coffee drinker": "Since my spoon is dry I must not have sugared my coffee, because the spoon would be wet if I had stirred the coffee, and I wouldn't have stirred it unless I had put sugar in it."

This is not a valid argument in the 2-valued logic, but there are still deductions and inferences to be drawn from these conditional premises. Let D denote "my spoon is dry"; let G denote "I sugared my coffee"; and let R denote "I stirred my coffee". Translating into this terminology the set of premises is $J = \{D, (D'|R), (R'|G')\}$. Therefore the conjunctive closure $C(J) = \{D, (D'|R), (R'|G'), D(D'|R), D(R'|G'), (D'|R)(R'|G'), D(D'|R)(R'|G')\}$. Using the operations on conditionals 1.1-1.4 C(J) becomes $\{D, (D'|R), (R'|G'), DR', DG \vee DR'G', (D'RG \vee R'G' \mid R \vee G'), DR'\}$. So according to the Corollary 2.2.12, for any of the elementary deductive relations $\leq_x$ or their combinations, except for $\leq_{tr}$ and $\leq_{\vee}$, $\mathcal{H}_x(J) = \mathcal{H}_x(D) \cup \mathcal{H}_x(D'|R) \cup \mathcal{H}_x(R'|G') \cup \mathcal{H}_x(DR') \cup \mathcal{H}_x(DG \vee DR'G') \cup \mathcal{H}_x(D'RG \vee R'G' \mid R \vee G')$.

Now this union can be simplified because some of these DCS's are included in the others. For instance, since all of these deductive relations satisfy $DR' \leq D$, therefore $\mathcal{H}_x(DR') \supseteq \mathcal{H}_x(D)$. Similarly, $DR' \leq DG \vee DR' = D(G \vee R') = D(G \vee R'G') = DG \vee DR'G'$. So $\mathcal{H}_x(DR') \supseteq \mathcal{H}_x(DG \vee DR'G')$. Thus, $\mathcal{H}_x(J) = \mathcal{H}_x(D'|R) \cup \mathcal{H}_x(R'|G') \cup \mathcal{H}_x(DR') \cup \mathcal{H}_x(D'RG \vee R'G' | R \vee G')$.

For x = ip, nf or $\wedge$, by Corollary 2.2.13, $\mathcal{H}_x(J) = \mathcal{H}_x(D(D'|R)(R'|G')) = \mathcal{H}_x(DR')$. Therefore $\mathcal{H}_{nf}(J) = \mathcal{H}_{nf}(DR') = \{(DR' \vee Y \mid Z)\text{: any } Y, Z \text{ in } \mathcal{B}\}$. That is, the implications of J when its conditionals are regarded as non-false, are all those conditionals with any condition and whose conclusion includes the event DR', that "my spoon is dry" and "I did not stir my coffee". Notice that G', "I did not sugar my coffee", is not an implication of J with respect to the non-falsity deductive relation, and neither is it a valid consequence of J in the 2-valued Boolean logic. In the 2-valued logic the implications of J are the universally conditioned events that include DR', that the spoon is dry and my coffee is not stirred. But the implications with respect to the "non-falsity" deductive relation $\leq_{nf}$ include all those with any other condition attached.

Similarly, by Corollary 2.2.13, $\mathcal{H}_{\wedge}(J) = \mathcal{H}_{\wedge}(DR') = \{(DR' \vee Y \mid Z)\text{: any } Y, Z \text{ in } \mathcal{B}\} = \mathcal{H}_{nf}(J)$, and so in this case the implications with respect to $\leq_{\wedge}$ are equal to the implications with respect to $\leq_{nf}$.

Turning to $\leq_{pm}$, there is an additional simplification. Since $DR' \leq_{pm} (D'|R)$, therefore $\mathcal{H}_{pm}(DR') \supseteq \mathcal{H}_{pm}(D'|R)$. So $\mathcal{H}_{pm}(J) = \mathcal{H}_{pm}(R'|G') \cup \mathcal{H}_{pm}(DR') \cup \mathcal{H}_{pm}(D'RG \vee R'G' \mid R \vee G')$.

By equation 2.2.6, $\mathcal{H}_{pm}(R'|G') = \{(R'G' \vee G \vee Y \mid R'G' \vee Z)\text{: any } Y, Z \text{ in } \mathcal{B}\}$, and $\mathcal{H}_{pm}(DR') = \{(DR' \vee Y \mid DR' \vee Z)\text{: any } Y, Z \text{ in } \mathcal{B}\}$, and $\mathcal{H}_{pm}(D'RG \vee R'G' \mid R \vee G') = \{(D'RG \vee R'G' \vee R'G \vee Y \mid D'RG \vee R'G' \vee Z)\text{: any } Y, Z \text{ in } \mathcal{B}\} = \{(D'RG \vee R' \vee Y \mid D'RG \vee R'G' \vee Z)\text{: any } Y, Z \text{ in } \mathcal{B}\}$.

So the set of implications with respect to $\leq_{pm}$ of $J = \{D, (D' \mid R), (R'|G')\}$ is the union of three sets of conditionals. $\mathcal{H}_{pm}(R'|G')$ is the set of all conditionals whose condition includes my not stirring nor sugaring my coffee and whose conclusion includes sugaring my coffee or not sugaring nor stirring it. $\mathcal{H}_{pm}(DR')$ is the set of all conditionals whose condition and conclusion include my not stirring my coffee and my spoon being dry. $\mathcal{H}_{pm}(D'RG \vee R'G' \mid R \vee G')$ is all conditionals whose condition includes my stirring and sugaring my coffee or not sugaring my coffee and whose conclusion includes my stirring and sugaring my coffee and wetting my spoon or neither stirring nor sugaring my coffee.

All the conditionals in $\mathcal{H}_{pm}(R'|G')$ have conditional probability no less than $P(R'|G')$. Those in $\mathcal{H}_{pm}(DR')$ have conditional probability no less than $P(DR')$, and those conditionals in $\mathcal{H}_{pm}(D'RG \vee R'G' \mid R \vee G')$ have conditional probability no less than $P(D'RG \vee R'G' \mid R \vee G')$.

If the spoon is observed to be dry, then D=1. So $\mathcal{H}_{pm}(DR') = \{(R' \vee Y \mid R' \vee Z)$: any Y, Z in $\mathcal{B}\}$ and $\mathcal{H}_{pm}(D'RG \vee R'G' \mid R \vee G') = \{(R' \vee Y \mid R'G' \vee Z)$: any Y, Z in $\mathcal{B}\}$, and the latter set of conditionals includes those of $\mathcal{H}_{pm}(DR')$ by setting $Z = R'G \vee W$, where W is any proposition or event in $\mathcal{B}$. Furthermore, the set $\mathcal{H}_{pm}(R'|G') = \{(R'G' \vee G \vee Y \mid R'G' \vee Z)$: any Y, Z in $\mathcal{B}\}$ $= \{(R' \vee G \vee Y \mid R'G' \vee Z)$: any Y, Z in $\mathcal{B}\}$ is also a subset of $\{(R' \vee Y \mid R'G' \vee Z)$: any Y, Z in $\mathcal{B}\}$ by setting $Y = G \vee W$, where W is any proposition or event in $\mathcal{B}$.

So if my spoon is observed to be dry (D=1), then $\mathcal{H}_{pm}(J) = \{(R' \vee Y \mid R'G' \vee Z)$: any Y, Z in $\mathcal{B}\}$. Thus the only conditionals implied with respect to $\leq_{pm}$ by $J = \{D, (D'|R), (R'|G')\} = \{1, (0|R), (R'|G')\}$ are $\{(R' \vee Y \mid R'G' \vee Z)$: any Y, Z in $\mathcal{B}\}$, namely those whose condition includes the non-stirring and non-sugaring of my coffee and whose conclusion includes the non-stirring of my coffee.

Finally consider the implications with respect to the deductive relation $\leq_{bo}$. From $\mathcal{H}_x(J) = \mathcal{H}_x(D'|R) \cup \mathcal{H}_x(R'|G') \cup \mathcal{H}_x(DR') \cup \mathcal{H}_x(D'RG \vee R'G' \mid R \vee G')$ it follows that $\mathcal{H}_{bo}(J) = \mathcal{H}_{bo}(D'|R) \cup \mathcal{H}_{bo}(R'|G') \cup \mathcal{H}_{bo}(DR') \cup \mathcal{H}_{bo}(D'RG \vee R'G' \mid R \vee G') = \{(D'R \vee Y \mid R)$: and Y in $\mathcal{B}\} \cup \{(R'G' \vee Y \mid G')$: any Y in $\mathcal{B}\} \cup \{(DR' \vee Y)$: any Y in $\mathcal{B}\} \cup \{(D'RG \vee R'G' \vee Y \mid R \vee G')$: any Y in $\mathcal{B}\}$.

2.3.3. Absent-minded Coffee Drinker Revisited. It interesting to see what happens with this example when the conditional (R'|G') in J is replaced by (R|G). Instead of saying "I wouldn't have stirred my coffee unless I had put sugar in it" suppose it was "if I sugared my coffee then I stirred it." Thus $J = \{D, (D'|R), (R|G)\}$.

In the Boolean 2-valued logic, the implications of J are those of the conjunction $D(D'|R)(R|G)$ where the conditionals are equated to their material conditionals and have a conjunction $D(D' \vee R')(R \vee G') = DR'G'$.

More generally the conjunctive closure of J is $C(J) = \{D, (D'|R), (R|G), D(D'|R), D(R|G), (D'|R)(R|G), D(D'|R)(R|G)\} = \{D, (D'|R), (R|G), DR', DG' \vee DRG, (D'RG' \vee D'RG \mid R \vee G), DR'G'\}$. Obviously, the propositions D and DR' are implications with respect to all deductive relations of DR'G', and so for all deductive relations $\leq_x$ their implications are included in $\mathcal{H}_x(J) = \mathcal{H}_x(D'|R) \cup \mathcal{H}_x(R|G) \cup \mathcal{H}_x(DG' \vee DRG) \cup \mathcal{H}_x(D'R \mid R \vee G) \cup \mathcal{H}_x(DR'G')$. Furthermore, $(DG' \vee DRG) = D(G' \vee RG) = D(G' \vee R) = D(R'G' \vee R) = DR'G' \vee R)$ is also an implication of DR'G'. So dropping $\mathcal{H}_x(DG' \vee DRG)$ from the union, $\mathcal{H}_x(J) = \mathcal{H}_x(D'|R) \cup \mathcal{H}_x(R|G) \cup \mathcal{H}_x(D'R \mid R \vee G) \cup \mathcal{H}_x(DR'G')$.

Note that the proposition DR'G' (having a dry spoon, unstirred coffee, and unsugared coffee) which is the conjunction of the three original conditionals of $J = \{D, (D' \mid R), (R|G)\}$, is an implication with respect to all these deductive relations. It is a logical consequence of J. Furthermore, by rearranging the conditioning, its probability $P(DR'G') = P(D)P(R'G'|D) = P(D)P((G'|R') \mid D)P(R'|D) = P(D)P(G'|DR')P(R'|D)$. This latter product has easily estimated conditionals probabilities. $P(D) = 1$ by observation, and both $P(G'|DR')$ and $P(R'|D)$ are also close to or equal to 1. This is one way the reasoning can proceed even though the initial phrasing was in terms of conditionals whose probabilities are not so easily estimated.

In addition, (D'R | R $\vee$ G) $\leq_{pm}$ (D'|R) because (D'R)(R $\vee$ G) $\leq$ (D'R) and D'R $\vee$ R'G' $\leq$ D'R $\vee$ R'. Thus $\mathcal{H}_{pm}$(J) = $\mathcal{H}_{pm}$(R|G) Ê $\mathcal{H}_{pm}$(D'R | R $\vee$ G) Ê $\mathcal{H}_{pm}$(DR'G'). So $\mathcal{H}_{pm}$(J) = {(RG $\vee$ G' $\vee$ Y | RG $\vee$ Z): any Y, Z in $\mathcal{B}$} Ê {(D'R $\vee$ R'G' $\vee$ Y | D'R $\vee$ Z): any Y, Z in $\mathcal{B}$} Ê {(DR'G' $\vee$ Y | DR'G' $\vee$ Z): any Y, Z in $\mathcal{B}$}. Furthermore, the conditionals in $\mathcal{H}_{pm}$(R|G) all have conditional probability no less than P(R|G), and similarly for the conditionals in $\mathcal{H}_{pm}$(D'R | R $\vee$ G) and in $\mathcal{H}_{pm}$(DR'G').

Turning to the non-falsity deductive relation, because DR'G' $\leq$ D'R $\vee$ R'G' therefore DR'G' $\leq_{nf}$ (D'R | R $\vee$ G), and so $\mathcal{H}_{nf}$(D'R | R $\vee$ G) $\subseteq$ $\mathcal{H}_{nf}$(DR'G'). Furthermore, because DR'G' $\leq$ R $\vee$ R'G' = R $\vee$ G' therefore DR'G' $\leq_{nf}$ (R|G) and so $\mathcal{H}_{nf}$(R|G) $\subseteq$ $\mathcal{H}_{nf}$(DR'G'). So the implications of J with respect to $\leq_{nf}$ are $\mathcal{H}_{nf}$(J) = {(DR'G' $\vee$ Y | Z): any Y, Z in $\mathcal{B}$}, namely any conditionals whose conclusion includes DR'G'.

Finally, with respect to $\leq_{bo}$ from $\mathcal{H}_{x}$(J) = $\mathcal{H}_{x}$(D'|R) Ê $\mathcal{H}_{x}$(R|G) Ê $\mathcal{H}_{x}$(D'R | R $\vee$ G) Ê $\mathcal{H}_{x}$(DR'G') it follows that $\mathcal{H}_{bo}$(J) = $\mathcal{H}_{bo}$(D'|R) Ê $\mathcal{H}_{bo}$(R|G) Ê $\mathcal{H}_{bo}$(D'R | R $\vee$ G) Ê $\mathcal{H}_{bo}$(DR'G') = {(D'R $\vee$ Y | R): any Y in $\mathcal{B}$} Ê {(RG $\vee$ Y | G): any Y in $\mathcal{B}$} Ê {(D'R $\vee$ Y | R $\vee$ G): any Y in $\mathcal{B}$} Ê {(DR'G' $\vee$ Y): any Y in $\mathcal{B}$}. So the implications with respect to $\leq_{bo}$ include $\mathcal{H}_{bo}$(D'|R), all those conditionals with the condition that I stirred my coffee (R) and with a conclusion that includes a non-dry spoon and stirred coffee (D'R). $\mathcal{H}_{bo}$(R|G) is all conditionals with sugared coffee (G) as condition and with a conclusion that includes RG, stirred and sugared coffee. $\mathcal{H}_{bo}$(D'R | R $\vee$ G) is all conditionals with conclusions that include D'R and with condition R $\vee$ G, of either stirred coffee or sugared coffee. $\mathcal{H}_{bo}$(DR'G') is simply the set of all (universally unconditioned) events that include DR'G', a dry spoon and unstirred, unsugared coffee. All of these conditionals have probabilities no less than the corresponding conditional that generates them.

3. Computations with Conditionals. While the preceding sections provide an adequate theoretical basis for calculating and reasoning with conditional propositions or conditional events, the problem of the complexity of information is no less daunting. Indecd, even without the added computational burden of operating with explicit conditionals, just operating with Boolean expressions in practical situations with, say, a dozen variables, is already too complex for practical pure Bayesian analysis. The reason for this is that in most situations the available information is insufficient to determine a single probability distribution that satisfies the known constraints of the situation. Various possibilities concerning unknown dependences between subsets of variables result in complicated solutions to relatively simple problems.

3.1 Pure Bayesian Analysis. For example, consider again the transitivity problem of Section 2.3.1. If "A given B" and "B given C" are both certain, then it follows that "A given C" is also a certainty. But if they are not certain, then by pure Bayesian analysis, P(A|C) can be zero no matter how high are the conditional probabilities of (A|B) and (B|C). This happens because P(B|C) and P(A|B) can be almost 1 while P(A| BC) is zero, and it is the latter probability that appears in the Bayesian solution: P(A|C) = P(AB or AB' | C) = P(AB|C) + P(AB'|C) = P(ABC)/P(C) + P(AB'C)/P(C) = P(ABC | BC) P(BC|C) + P(AB'C | B'C) P(B'C|C) = P(A|BC)P(B|C) + P(A|B'C)P(B'|C). Without knowing anything about P(A|BC) or P(A|B'C), nothing more can be said about P(A|C).

3.2 Choosing a Bayesian Solution Using Maximum Information Entropy. Continuing the example of Section 3.1, knowing that C is true might dramatically change P(A|B) up or down. But if nothing is known one way or the other, the choice of the maximum information entropy distribution assumes that P(A|BC) = P(A|B). This latter equation is called the *conditional independence* of A and C given B. It can also be expressed as P(AC|B) = P(A|B)P(C|B) or as P(C|AB) = P(C|B). Using this principle P(A|C) = P(A|B)P(B|C) + P(A|B'C)P(B'|C). So if P(A|B) and P(B|C) are 0.9 and 0.8 respectively, then P(A|C) is at least 0.72. Additionally, since nothing is known one way or the other about the occurrence of A when B is false and C is true, this principle of "maximum indifference" implies that P(A|B'C) should be taken to be ½. So the term P(A|B'C)P(B'|C) contributes (1/2)P(B'|C) = (1/2)(1 – 0.8) = 0.1 to P(A|C) bringing the total to 0.82.

In affect the principle of maximum information entropy chooses that probability distribution P that assumes conditional independence of any two variables that are not explicitly known to have some dependence under the condition. This greatly simplifies computations and often allows situations of several dozen variables to be rapidly analyzed as long as the clusters of dependent variables are not too large and not too numerous. The maximum entropy solution is always one of the possible Bayesian solutions of the situation. If there is just one Bayesian solution, then the two solutions will always agree.

It is a remarkable fact that such a function as the entropy function exists, and it is now clear that it has wide application to information processing under uncertainty. If the n outcomes of some experiment are to be assigned probabilities p_i for i=1 to n subject to some set of constraints, then the distribution of probabilities that assumes conditional independence unless dependence is explicitly known is the one that maximizes the entropy function

$$H(p_1, p_2, p_3, \ldots, p_n) = - \sum_{i=1}^{n} p_i \log p_i$$

and also satisfies the known constraints. If there is an a priori distribution $q_1, q_2, q_3, \ldots, q_n$ then H is given by

$$H(p_1, p_2, p_3, \ldots, p_n, q_1, q_2, q_3, \ldots, q_n) = - \sum_{i=1}^{n} p_i \log (p_i/q_i)$$

This allows maximum entropy updates when additional information is available. See J. E. Shore [Sho80] for a derivation.

W. Rödder [Röd96, Röd00] and his colleagues at Fern University in Hagen are continuing to develop a very impressive interactive computer program SPIRIT that implements this practical approach to the computation of propositions and conditional propositions and their probabilities. Starting with an initially defined set of variables and their values, the user can input statements and conditionals statements about these variables taking various values, and can also assign conditional probabilities to them. The *utility* of having a variable take one of its values can also be incorporated.

3.3 Confidence in Maximum Entropy Solutions. While the maximum entropy solution provides the most plausible or "most likely" probability distribution for a situation among all of the Bayesian solutions, it does not immediately provide a means for estimating how much confidence to attach to that solution. This issue has been taken up by E. Jaynes [Jay79], S. Amari [Ama85]], and A. Caticha [Cat00].

Jaynes puts the matter as follows: "Granted that the distribution of maximum entropy has a favored status, in exactly what sense, and how strongly, are alternative distributions of lower entropy ruled out?" He proves an entropy "concentration theorem" in the context of an generalized experiment of N independent trials each having n possible results and satisfying a set of m (< n) linearly independent, linear constraints on the observed frequencies of the experiment. Jaynes shows that in the limit as the number N of trials approaches infinity, the fraction F of probability distributions satisfying the m constraints and whose entropy H differs from the maximum by no more than ΔH is given by the Chi-square distribution χ_k^2 with $k = n - m - 1$ degrees of freedom as

$$2N(\Delta H) = \chi_k^2(F)$$

That is, the critical, threshold entropy value H_α for which only the fraction α of the probability distributions that satisfy the m constraints have smaller entropy is given by

$$H_\alpha = H_{max} - \chi_k^2(1-\alpha) / 2N.$$

For N = 1000 independent trials of tossing a 6-sided die and with a significance level $\alpha = 0.05$ and degrees of freedom k = 6 - 1 - 1 = 4, 95% of the eligible probability distributions have entropy no less than $H_\alpha = H_{max} - 9.49 / 2N = H_{max} - 0.0047$. H_{max} is on the order of 1.7918 for a fair die and 1.6136 for a die with average die value of 4.5 instead of 3.5. Letting $\alpha = 0.005$ it follows that 99.5% of the eligible distributions will have entropy no less than $H_{max} - 14.9 / 2000 = H_{max} - 0.00745$.

Clearly eligible distributions that significantly deviate in entropy from the maximum value are very rare. However this result does not directly answer the question of how much confidence to have in the individual probabilities associated with distributions having maximum or almost maximum entropy. That is, can a probability distribution with close to maximum entropy assign probabilities that are significantly different from the probabilities of the maximum entropy distribution?

For instance, a 6-sided die having two faces with probabilities 1/12 and 1/4 respectively and four faces each having 1/6 probability has entropy 0.0436 less than the maximum of 1.7918 for a fair die. So for N=1000 independent trials and a significance level of $\alpha = 0.05$ such a distribution would differ from the maximum entropy value for a fair die by considerably more than 0.0047. However for N=100, $\Delta H = 9.49/200 = 0.047$, which is large enough to include such a distribution.

Furthermore, how does the confidence in the probabilities determined by a maximum entropy solution depend upon the amount of under-specification of the situation that produced that solution? Surely a maximum entropy distribution that relies upon a great deal of ignorance about a situation offers less confidence about the probabilities determined than does a maximum entropy solution that is based upon a minimum of ignorance about the situation. Put another way, confidence about the maximum entropy distribution should be higher when conditional independencies are positively known than when they are merely provisionally assumed.

Amari [Ama85] takes up these issues in the context of differential geometry. Under-specification of information gives rise to a manifold of possible probability distributions. A Riemannian metric on these distributions early introduced by C. R. Rao [Rao45] allows a very general approach to quantifying the distance between distributions. This development provides a very general approach to these problems of multiple possible distributions, but so far the results don't seem to directly apply to the issue of the confidence to be attached to the individual probabilities dictated by a maximum entropy distribution. Unfortunately Amari offers no numerical example to illustrate how these results might be applied to allow a confidence measure to be put upon the probabilities associated with distributions having maximum or close to maximum entropy.

Caticha [Cat00] frames the question along the same lines as Jaynes: "Once one accepts that the maximum entropy distribution is to be preferred over all others, the question is to what extent are distributions with lower entropy supposed to be ruled out?" Using a parameterized family of distributions Caticha shows how this question can be rephrased as another maximum entropy problem, but he too offers no simple illustrative example of how his results can be applied to the question of how much confidence to have in any one probability value associated with the maximum entropy distribution.

What seems to be needed is a way to solve for the probabilities of specified outcomes in terms of entropies equal to or close to the maximum entropy. If 95% of the eligible probability distributions have entropy H no less than $H_{max} - \Delta H$, then what confidence limits are implied for the individual probabilities of those distributions?

4. **Summary.** In order to adequately represent and manipulate explicitly conditional statements such as "A given B" the familiar Boolean algebra of propositions or events must be extended to ordered pairs of such propositions or events. This is quite analogous to the requirement to extend integers to order pairs in order to adequately represent fractions and allow division. The resulting system of Boolean fractions includes the original propositions and also allows the non-trivial assignment of conditional probabilities to these Boolean fractions. Boolean fractions are truth functional in the sense that their truth status is completely determined by the truth or falsity of the two Boolean components of the fraction. But since there are two components, the truth status of a Boolean fraction has three possibilities – one when the condition (denominator) is false and two more when the denominator is true. Just as all integer fractions with a zero denominator are "undefined", so too are all Boolean fractions with a false condition undefined or "inapplicable". When the condition is true then the truth status of a Boolean fraction is determined by the truth of the numerator. The four extended operations (or, and, not, and given) on the Boolean fractions reduce to ordinary Boolean operations when the denominators are equivalent. Just as with integer fractions, the system of Boolean fractions has some new properties but loses others that are true in the Boolean algebra of propositions or events.

A conditional statement is not an implication or a deduction; it is rather a statement in a given context. Deduction of one conditional by another can still be defined in terms of the (extended) operations, as is often done in Boolean algebra. Due to the two components of a conditional there is a question of what is being implied when one conditional implies another. It turns out that several plausible implications between conditionals can be reduced to ordinary implications between the Boolean components of the two conditionals. The applicability, truth, non-falsity or inapplicability of one conditional can imply the corresponding property in the second conditional. Any two or more of these four elementary implications can be combined to form a more stringent implication. With respect to any one of these implications, a set of conditionals will generally imply a larger set, and it is now possible to compute the set of all deductions generated by some initial set of conditionals, as illustrated by three examples in this paper.

While computations can be done in principle, in practice the complexity of partial and uncertain conditional information precludes the possibility of solving for all possible probability distributions that satisfy the partial constraints. What is feasible and already successfully implemented in the program SPIRIT is to compute the distribution with maximum information entropy. However, the amount of confidence that can be associated with the probabilities assigned by this "most likely", maximum entropy distribution is still an open question.

References & Bibliography

[Ada66] E. W. Adams, "Probability and the Logic of Conditionals," in: *Aspects of InductiveLogic*, P. Suppes & J. Hintikka, eds., Amsterdam: North-Holland (1966) 265-316.

[Ada86] E. W. Adams, "On the Logic of High Probability", *Journal of Philosophical Logic* 15 (1986) 225-279.

[Ada98] E. W. Adams, A primer of Probability Logic, *CSLI Publications* (1998).

[Ama85] S. Amari, *Differential-Geometrical Methods in Statistics*, Springer-Verlag, 1985.

[Ama89] S. Amari & T.S. Han, "Statistical Inference Under Multiterminal Rate Restrictions: A Differential Geometric Approach", Information Theory, *IEEE Transactions on*, Volume: 35 Issue: 2, (March 1989) 217 –227

[Cal87] P. G. Calabrese, An Algebraic Synthesis of the Foundations of Logic and Probability, *Information Sciences*, 42, (1987) 187-237.

[Cal90] P. G. Calabrese, "Reasoning with Uncertainty Using Conditional Logic and Probability," in: *Proceeding of the First International Symposium on Uncertainty Modeling and Analysis*, IEEE Computer Society (1990) 682-688.

[Cal91] P. G. Calabrese, Deduction and Inference Using Conditional Logic and Probability, in: I. R. Goodman, M. M. Gupta, H. T. Nguyen and G. S. Rogers, eds., *Conditional Logic in Expert Systems*, North-Holland, Amsterdam (1991) 71-100.

[Cal93] P. G. Calabrese, An extension of the fundamental theorem of Boolean algebra to conditional propositions, Part I (1-18) of "Conditional event algebras and conditional probability logics" by P. G. Calabrese and I. R. Goodman, *Probabilistic Methods in Expert Systems*, Romano Scozzafava ed., Proceedings of the International Workshop, Rome, Italy, Societá Italiana di Statistica, (1993) 1-35.

[Cal94] P. G. Calabrese, "A Theory of Conditional Information with Applications", In *Special Issue on Conditional Event Algebra, IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 24, No. 12 (1994) 1676-1684.

[Cal02] P. G. Calabrese, "Deduction with Uncertain Conditionals", accepted for publication in *Information Sciences*.

[Cat00] A. Caticha, "Maximum Entropy, Fluctuations and Priors", MaxEnt 2000, the 20th International Workshop on Bayesian Inference and Maximum Entropy Methods, July 8-13, 2000, Gif-sur-Yvette, France.

[DeF36] B. De Finetti, "La logique de la probabilite," in *Actes du Congres International de Philosophie Scientifique* Paris: Hermann Editeurs, IV, (1936) 1-8.

[DeF72] B. De Finetti, *Probability, Induction and Statistics*, Wiley, New York (1972).

[Dom69] Z. Domotor "Probabilistic Relation Structures and their Applications" (Technical Report 144), Stanford University, Institute for Mathematical Studies in the Social Sciences, Stanford, California.

[Dub94] D. Dubois, and H. Prade, "Conditional Objects as Nonmonotonic Consequence Relationships", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 24, No. 12 (1994) 1724-1740.

[Goo91A] I. R. Goodman, H. T. Nguyen, E. A. Walker, *Conditional Inference and Logic for Intelligent Systems: A Theory of Measure-Free Conditioning*, North-Holland, 1991.

[Hai96] T. Hailperin, *Sentential Probability Logic*, Lehigh Univ. Press (1996).

[Jay79] E.T. Jaynes, "Concentration of Distributions at Entropy Maxima", reprinted in R. D. Rosenkrantz (ed.), *E.T. Jaynes: Papers on Probability, Statistics and Statistical Physics*, Reidel, Dordrecht, 1983.

[Lew76] D. Lewis, "Probabilities of Conditionals and Conditional Probabilities", *The Philos. Rev.* 85(3) (1976), 297-315.

[Pos71] H. Pospesel, *Arguments: Deductive Logic Exercises*, Prentice Hall (1971).

[Res69] N. Rescher, *Many-Valued Logics*, McGraw-Hill (1969).

[Röd96] W. Rödder, C.-H. Meyer, "Coherent Knowledge Processing at Maximum Entropy by SPIRIT", Proc. Twelfth Conf. On Uncertainty in Artificial Intelligence, Portland, Oregon, USA, (1996) 470-476.

[Röd00] W. Rödder, "Conditional logic and the Principle of Entropy", *Artificial Intelligence* 117 (Feb. 2000) 83-106

[Rao45] C. R. Rao, "Information and the Accuracy Attainable in the Estimation of Statistical Parameters", *Bulletin Calcutta Mathematical Society*, Vol. 35, (1945) 199-210.

[Sch68] G. Schay, "An Algebra of Conditional Events", *Journal of Mathematical Analysis and Applications 24* (1968), 334-344.

[Sho80] J. E. Shore, "Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy", *IEEE Transactions on Information Theory*, IT-26, 1 (Jan. 1980), 26-37.

[Sob52] B. Sobocinski, "Axiomatization of a partial system of the three-valued calculus of propositions", *J. Computing Systems* 1 (1) 23-55.

[Wal94] E. A. Walker, "Stone Algebras, Conditional Events, and Three valued Logic", *IEEE Transactions of Systems, Man and Cybernetics*, Vol. 24, Number 12 (Dec. 1994) 1699-1707.